64-9

429434

EFFECTIVE INDIVIDUAL PERFORMANCE IN SMALL ANTARCTIC STATIONS: A SUMMARY OF CRITERION STUDIES

PAUL D. NELSON

E. K. ERIC GUNDERSON

REPORT NUMBER: 63-8

U. S. NAVY MEDICAL
NEUROPSYCHIATRIC RESEARCH UNIT
SAN DIEGO 52, CALIFORNIA

BUREAU OF MEDICINE AND SURGERY NAVY DEPARTMENT
WASHINGTON 25, D. C.

EFFECTIVE INDIVIDUAL PERFORMANCE IN SMALL ANTARCTIC STATIONS: A SUMMARY OF CRITERION STUDIES

by

Paul D. Nelson
Lieutenant, MSC, U. S. Navy

and

E. K. Eric Gunderson, Ph.D.

United States Navy Medical Neuropsychiatric Research Unit

San Diego 52, California

Report Number 63-8, April 1963, supported by the Bureau of Medicine and Surgery, Department of The Navy, under Research Task MR005.12-2004, Subtask 1.

INTRODUCTION

Ambiguities which often arise with regard to the criterion concept are probably a function of the many ways in which psychologists define and use criteria. On a most general level of analysis, a criterion is a form or an outcome of behavior, operationally specified in advance as the goal of a behavior sequence, often as an object of prediction. The behavioral content of a criterion can be of a neurophysiological, attitudinal, or psychomotor nature. A criterion may refer to a completed act, an ongoing state of the organism, or perhaps the time and energy expended by the organism in carrying out some behavior.

Whatever the specific nature of a criterion may be, its selection depends upon the purpose of the investigator or the persons for whom he is a consultant. In some instances the psychologist alone can define his criterion, an example being the selection of an empirical referent against which to validate a previously established psychological test or construct. On the other hand, there are times when persons other than the psychologist may define the criterion appropriate for their situational or institutional goals, examples being criteria of job performance in industrial and military settings. Ultimately, any criterion must be meaningful and valid for the life situation in which it is used and for the persons involved in its use.

THE PRESENT PROBLEM

The present paper represents a summary of the research conducted for the purpose of developing a criterion of effective individual performance among persons who winter-over at small Antarctic scientific stations.

During each of the past six years, from 15 to 40 men have worked and lived at each of several Antarctic stations for 12 continuous months. For at least six months of each year the men have no way of leaving the station and, except for occasional radio communication, have no contact with the outside world. At most stations approximately half the personnel are Navy enlisted men whose primary responsibility is that of maintaining and operating the station itself; the remaining station members are civilian scientists and technicians whose primary responsibility is the collection of scientific data. In addition to their task specialties, all station members are dependent upon one another for the solution of common problems ranging from housekeeping chores to survival.

The need for the present criterion research resulted from a requirement that there be a psychiatric screening and selection program for persons volunteering to winter-over in the Antarctic. Any personnel selection program operates on a basic assumption that there is a preferred or expected form of behavior in the situation of concern and furthermore that not all persons have equal potential for satisfying such behavioral expectations. In most situations, for example, at least some minimum job skill level is required. But when a psychiatric screening and selection program is required there is an additional assumption that while an individual's technical competence may be a necessary attribute it is not in itself a sufficient qualification.

Since it is impossible to evacuate personnel from the stations during the Antarctic winter months, a major criterion at the time of screening is that the individual appear capable of a normal range of physical and psychological functioning; those judged to have any disruptive potential are screened out. But for the vast majority of volunteers, men who are initially acceptable on a

physical, psychiatric, and job skill basis, the problem remains to select those individuals with the highest potential for effective performance at a small station. What, then, constitutes effective performance and, if there are individual differences in such behavior, how can effective performance best be measured?

DEFINING A CRITERION

An Overview

A basic assumption with regard to any criterion of effective performance in Antarctic stations is that there are individual differences. Fortunately, survival itself has not served as an adequate criterion on which to differentiate persons nor has the incidence of gross physical or psychiatric malfunctioning (Nardini, Herrmann, and Rasmussen, 1962). In terms of task fulfillment, it has been difficult to assess the extent to which individuals have accomplished predetermined goals. For one thing, it is not always known in advance what every man will have to accomplish in his work except in a broad sense; then too, unpredicted events beyond the individual's control can disrupt his achievement of established goals; and finally, with the different types of tasks to be performed, it has been difficult to establish a meaningful standard of output applicable to all personnel.

In the absence of the preceding types of criteria, and in view of the small station parameters, effective individual performance has been defined in terms of the following three components: emotional composure, ability to get along with others and work efforts (Nelson, 1962). On face value the extent to which an individual possesses positive attributes in each of these three behavior domains should determine the extent to which his performance can be judged effective. While it is recognized that performance is a continuous and undoubtedly fluctuating process, the search for a criterion measure in the present context will be characterized by the assumption that the most meaningful criterion should be one reflecting the individual's general behavior over the entire Antarctic year.

The only available source of information about the individual's performance over the entire year is the station membership itself. It furthermore seems quite reasonable to assume that the station leaders or supervisors and the other station members (peer group) are as capable as anyone in observing whether or not an individual becomes emotionally distressed, socially disruptive, or is lazy or incompetent. In other industrial and military settings, supervisor evaluations are routine; and, in recent years, much evidence has been gathered indicating the value of peer evaluations in operational field settings (Anderhalter, Wilkins, and Rigby, 1952; Berkshire and Nelson, 1958; Hollander, 1954; Wherry and Fryer, 1949).

Earlier Research

During the first three years of Antarctic station activity, attempts were made to collect supervisor and peer evaluations of station members' overall performance and of their emotional, social, and work characteristics. Much of these data were incomplete; for some stations there were no data collected at all; and the specific items on which evaluations were made varied from year to year. Analyses of those data which were obtained (Gunderson and Nelson, 1962 (b); Nelson and Gunderson, 1962) provided the following general results: (1) significant agreement was found between supervisors' and peers' evaluations both within and between time periods and (2) significant positive intercorrelations were found among evaluations of social, work, and overall performance.

The implications of these earlier results are that station members can be reliably differentiated by peers and supervisors on the basis of their performance, that individual performance tends to be consistent over time periods, and that overall evaluations of performance reflect both the social and work qualities of the individual. While less information was collected on emotional composure, it has been assumed that effective emotional control contributes to effective social and work performance.

Recent Research

During the past three years somewhat different assessment techniques have been tried taking advantage of knowledge gained from earlier attempts to collect data in this particular field setting. First of all, for obtaining evaluative information in the Antarctic, the Navy officer and civilian supervisor at each station are usually asked to provide more information than the peer group members; past experience has indicated more difficulty in obtaining personnel evaluations from peer group members than from supervisors during the Antarctic year itself. When peer evaluations are obtained in the Antarctic, they are now obtained in the form of positive nominations only. During the most recent year in which the positive nomination technique was used, approximately 76 per cent of all small station members gave complete information with a higher percentage giving at least partial information. This represented the most successful year to date for obtaining peer evaluations in the Antarctic. In addition, to supplement peer evaluations obtained in the Antarctic, it has been found feasible to obtain valid evaluations from peer group members by mail questionnaire once they have returned from the Antarctic and up to one year later (Nelson, 1963). In summary, using each of the preceding methods, data are now available on the personnel from seven of nine small stations operated during the past three years (N = 139).

At the end of each Antarctic year, the two supervisors at each station (one Navy officer and one civilian supervisor) have independently evaluated all station personnel, using nine-point graphic rating scales, on the following characteristics or traits: emotional control, self-confidence, achievement motivation, acceptance of authority, likability by group members, alertness, industry, happiness, motivation to be an efficient group member, attitude towards the Antarctic project, and satisfaction with job assignment.[1] Also, each supervisor has ranked all station members in the order in which he would select them for wintering-over duty were he to return to an Antarctic station.

The peer group members from four of the seven stations during the past three years provided evaluations of one another through mail questionnaire upon return from the Antarctic. Each individual ranked his former station members, just as the supervisors did, in the order in which they would be selected for wintering-over duty. Each man also indicated his closest friends and, using an open-end form, described the strong and weak points of every station member. The peer group members from the other three stations provided evaluations at each of three times during the Antarctic year itself. Each man nominated five station members for each of the following: (1) easiest to get along with, (2) contributes most through his work efforts, and (3) would select to return with for wintering-over duty again.

---

[1] The items were selected from the Personal Adjustment Booklet (Modified Deep Freeze) prepared by Benjamin B. Wuybrew, Ph.D., USN, Medical Research Laboratory, New London, Connecticut. Several of the original items were not included in the present study due to any of several reasons, such as inability of supervisors to use them, their redundant nature relative to items presently included, or multi-dimensionality.

The one item common to all assessment forms during the past three years is that pertaining to the selection (by rank or nomination) of station members with whom one would prefer to return to the Antarctic were he to have such duty again. This item has been considered, on face value, to be the most meaningful single index of effective performance. On the basis of the cooperation obtained in using this particular item, it appears to be meaningful to the respondents themselves. This item, then, has served as the basis for the development of a criterion of effective individual performance.

DEVELOPING A CRITERION SCORE

On the basis of earlier research (Nelson and Gunderson, 1962) and on theoretical grounds, it seemed reasonable to assume that supervisors and peer group members may have somewhat different frames-of-reference in evaluating personnel. It may be, for example, that supervisors are more concerned than peers with getting the job done while peers are more concerned with interpersonal relations. This being possible, each station member was given a criterion score based upon peer evaluations (peer criterion) and one based upon supervisor evaluations (supervisor criterion), using the "return with" item as the basis for such scores. If there were any differences in standards of judgment, then, they would be reflected in the two scores. The scores were developed in the following manner.

Within each station the members were ranked from 1 through N on the basis of (1) an average of the two supervisors' rankings and (2) an average of the peer group's rankings or nominations on the "return with" item. At the three stations from which nominations, rather than ranks, were obtained, nominations were weighted from +5 to +1 and individuals were given an average score across time periods; ranks were then determined from these average scores. All ranks were then converted to T-scores (Mean = 50, SD = 10) so as to allow for the pooling of individuals from stations of slightly different size.

If peers and supervisors are both considered as important sources for the evaluation of an individual's performance, it would seem that some combination of the peer and supervisor criteria scores would represent the best single index of effective performance. If the two groups of raters have identical frames-of-reference and observe all persons in the same way, either the peer or supervisor criterion score alone would probably be sufficient; but to the extent that differences exist in their standards of judgment, the best overall performance would seem to be that which is acceptable to both peers and supervisors.

Since there is no particular reason to believe that either the peer or supervisor criterion should be considered more important than the other, equal weight was given to each in combining the criterion scores. The combined criterion score was therefore derived by averaging the T-scores from the separate peer and supervisor criteria. Individuals were not re-ranked on the average T-scores, thus making it possible to differentiate individuals from different stations who may have been best or poorest in their respective groups but of slightly different calibre in comparison to one another. To receive a high or low score, then, an individual must have been regarded with equal favor or disfavor by both the peer and supervisor groups of his station.

Before evaluating the characteristics of the criterion scores, a few comments on the appropriateness of the rank-based score seem necessary. Ranking does provide a standard instructional set for persons making evaluations, thus avoiding the problem of individual differences in preference for different levels and ranges of ratings. The ranking system also maximizes discrimination between stimuli (individuals), a property which is important in the present criterion development.

On the other hand, if the individuals of one group are homogeneously better or poorer performers than those of another group, ranking within groups would tend to obliterate such differences. While there is evidence that station groups have differed from one another in attitudes (Gunderson and Nelson, 1962(b)), there has been no conclusive evidence to support the notion that individual performance has been homogeneously better in one group than in another. Furthermore, on the basis of what is known of the initial assignment process, there is no reason at present to believe that the calibre of men at one station has been better overall than that at another.

But suppose there were such differences between stations. At present there is no common standard by which this could be verified; the only evidence would be in the form of differences between stations in terms of average ratings were they to be used in preference to ranks. Differences between ratings could be as much a function of the different raters as they are a function of actual differences in the performance of individuals being evaluated (Fiske and Cox, 1960). Of course, as the number of raters increases, the preceding argument would seem to be less valid. In the present situation, the ranking of individuals in a homogeneously high or low performance group could conceivably result in an almost random assignment of ranks by each judge; the subsequent lack of variance in combined evaluations would result in appreciably low reliability. As reported in the next section of this paper, the reliabilities have not been appreciably low. In summary, we presently assume that individual differences are as great within as between stations and that such differences can be estimated by a ranking technique.

CHARACTERISTICS OF CRITERION SCORES

Table 1 shows the distribution of combined criterion scores for the total population of persons from seven stations during the past three years. A normal distribution is maintained with the average T-scores.

TABLE 1

Distribution of Combined Criterion Scores[a]

Criterion Score Interval	f	%
68 - 72	3	2.4
63 - 67	10	7.1
58 - 62	15	10.7
53 - 57	26	18.7
48 - 52	32	23.0
43 - 47	23	16.5
38 - 42	18	12.9
33 - 37	9	6.4
28 - 32	3	2.4

[a]Derived from an average of the peer and supervisor criterion T-scores.

Mean = 49.98, SD = 8.71, Median = 50, N = 139

TABLE 2

Rater Agreement in The Ranking of Station Members on The Criterion Item

	Station						
Rater Groups	1	2	3	4	5	6	7
Military-civilian supervisor agreement	.53	.54	.45	.50	.33	.39	.58
Split-half peer group agreement	.59	.76	.70	.84	.36	.55	.54
Combined peer-supervisor agreement	.71	.84	.63	.65	.31	.69	.57
N	19	17	18	19	33	20	14
Pearson r ($p < .05$)	.46	.48	.47	.47	.34	.44	.53
Pearson r ($p < .01$)	.58	.61	.59	.59	.44	.56	.66

Data estimating the agreement among peers and supervisors in their criterion evaluations are contained in Table 2. The actual reliabilities, if computed with the Spearman-Brown prophecy formula, would yield higher coefficients than those tabled. Agreement between supervisors was estimated by correlating (Pearson r) the T-scores corresponding to the ranks given group members by each of the two supervisors at each station. Peer agreement was obtained by the same correlation technique applied to split-halves of the peer group; an alphabetical split was used and an even number of civilian and military group members were contained in each half. For only the largest station was the agreement between peers and supervisors insignificant.

One further analysis of the criterion scores was that of comparing the average criterion scores of individuals from different occupational groups. It may have been that the perceived importance of an occupation was a determining factor in selection of persons to "return with." No significant difference was revealed between the average criterion scores of military and civilian personnel. And, except for the cooks (N = 7) who had a higher criterion average than all other groups, an analysis of variance revealed no significant differences between the specific occupational groups (i.e., radioman, builder, mechanic, meteorologist, etc.). Criterion evaluations, therefore, were not determined to any important degree by the occupation in which a man worked.

THE CONCURRENT CORRELATES OF CRITERIA

The final phase of the criterion analysis was that of determining the specific traits or characteristics associated with the criterion scores. That is, on the basis of what behavior characteristics exhibited in the Antarctic is an individual selected by his peers and supervisors for wintering-over duty again? It is on the basis of such specific components of the criterion that the selection of useful screening variables might be chosen. As mentioned initially, effective performance has been considered to be a function of emotional,

social, and work attributes; it was therefore anticipated that evaluations of an individual's behavior relevant to those three areas would correlate significantly with the combined criterion score. As with the criterion of "return with," the trait evaluations were also obtained from peers and supervisors.

Peer Descriptions

As previously mentioned, peer group members from four of the seven stations gave open-end descriptions of the strong and weak points of all former station members on a follow-up mail questionnaire. A content analysis was applied to these descriptions. Table 3 contains groupings of the most frequently given types of descriptions, positive and negative, with category headings which seemed appropriate. To estimate the discriminatory power of the descriptions, individuals from the four stations with upper and lower quintile criterion scores were compared with one another. The relative frequency with which the different types of descriptions were given to these two criterion groups is also shown in Table 3.

The "task competence" and "social orientation" categories were least discriminating of the various groups of descriptions. The low incidence of descriptions of task incompetence may be indicative of the initial selection of at least minimally qualified men in each occupation for duty at the small stations. The fact that "social orientation" (an extrovert-introvert type of category) did not discriminate between the top and bottom performers supports a previously reported hypothesis that being predominantly outgoing or withdrawn is less important than the extent to which an individual is considerate of others when in fact he is in personal contact with others (Nelson, 1962). The other categories of behavior discriminated upper from lower criterion groups in a direction which would be expected, assuming the importance of emotional, social, and work characteristics for overall performance.

In addition to the preceding analysis, all individuals from the four stations (N = 72) were given algebraic sum scores based upon the relative number of positive and negative descriptions received in each of the three general areas of emotional, social, and work performance. The positive and negative values of the descriptions were determined on the basis of whether the description was given as a strong or a weak point. The members of each station were then dichotomized on each of the three dimensions, as close to the median as possible, upper and lower halves were pooled across the four station groups, and biserial correlations were computed against the combined criterion score (peer-supervisor average). The correlations of emotional, social, and task description scores with the criterion were .35, .42, and .65 respectively; all values are significant at the .01 level of confidence.

Of all descriptions given, 48 percent were in the task area, 34 percent in the social area, and 17 percent in the emotional area of behavior. Both the importance of and greater familiarity with concepts related to work performance probably contributed to the greater saliency of task descriptions. In any event, the fact that emotional behavior was least correlated with the criterion may not be so much a function of its lesser importance but rather a consequence of the relatively low variance obtained in emotional description scores due to the relatively infrequent use of descriptions in that area.

Instead of the preceding types of data, the other three stations provided peer nominations (weighted +5 to +1) at three times during the year on 1) easy to get along with (social) and 2) contributes through his work efforts (task). After giving each individual an average nomination score for each characteristic based upon nominations from all three time periods, Pearson r's were obtained between social, task, and combined criterion scores within each station. Using Fisher's z transformation technique, the average correlations across stations were .29, .76, and .58 (SE = .13) for social with task, social with criterion, and task with criterion respectively. These data suggest some unique contributions to the criterion by the social and task nomination scores. Based upon

the above correlations, a multiple correlation between social and task with the criterion is approximately .85.

TABLE 3

Percentages of Positive and Negative Descriptions Given to Persons in Upper and Lower Quintile (20%) on Combined Criterion

	Descriptions	Total N of Descriptions Given	% Given to UQ (N = 16)	% Given to LQ (N = 16)	% Given to LQ (N = 16)
Task Motivation					
(+)	ambitious, conscientious, dedicated, energetic, helpful, industrious	119	68%	32%	32%
(-)	lazy, slow, uncooperative	31	13%	87%	87%
Task Competence					
(+)	competent, efficient, experienced, intelligent, inventive, knows job, reliable, versatile	100	47%	53%	53%
(-)	incompetent, inexperienced, unreliable, couldn't do job	5	40%	60%	60%
Social Orientation					
(+)	participates, outgoing, sociable, talkative	7	57%	43%	43%
(-)	doesn't participate, quiet, shy, withdrawn	16	81%	19%	19%
Social Competence					
(+)	compatible, avoids arguing, considerate, friendly manner, takes criticism, unassuming, tolerant	26	73%	27%	27%
(-)	antagonistic, boastful, coarse, critical, egotistical, loud, inconsiderate, intolerant	35	23%	77%	77%
Emotional Control					
(+)	calm, easy going, even-tempered	9	78%	22%	22%
(-)	excitable, quick-tempered	9	00	100%	100%
Emotional Disposition					
(+)	cheerful, good natured, sense of humor, doesn't complain	34	82%	18%	18%
(-)	complaining, irritable, moody, sensitive, worrying	17	18%	82%	82%

Supervisor Descriptions

The supervisors at each station, in addition to providing rank criterion data, rated all station members on behavior characteristics exhibited during the year. The traits or characteristics are shown in Table 4 along with their relationships with the criteria and estimates of supervisor agreement on each characteristic.

TABLE 4

Trait-Criteria Correlations[a]

Trait	Criteria: Peer	Supervisor	Combined	Reliability[b] Estimate
Likability	.66	.73	.78	.55
Emotional control	.49	.73	.68	.57
Accept authority	.38	.62	.56	.67
Industriousness	.40	.60	.56	.63
Achievement motivation	.37	.59	.53	.45
Motivation towards group	.35	.60	.53	.47
Attitude towards project	.27	.60	.48	.56
Happiness	.37	.50	.48	.51
Alertness	.26	.51	.43	.40
Job satisfaction	.15	.49	.35	.54
Self-confidence	.24	.28	.29	.23

[a] Correlations are Pearson r's based upon T-scores for trait-criteria relationships; for N = 139, r = .16 ($p < .05$).

[b] Pearson r's based upon median agreement between supervisors' ratings across stations.

Supervisor agreement on each characteristic was obtained by correlation of the ratings given by two supervisors at each station (Pearson r). The median correlation values (across stations) are shown in Table 4. Some characteristics had greater agreement than others, but except for self-confidence the levels of agreement were reasonably similar. By ranking the characteristics for each pair of supervisors in terms of amount of rating agreement, a coefficient of concordance ($W_c = .265$) revealed significantly similar rank orders for different pairs of supervisors ($p < .01$). Some characteristics, therefore, may be consistently more difficult to observe or perhaps ambiguous to the supervisors.

Prior to obtaining the trait-criterion correlations, the supervisors' ratings were averaged on each characteristic and individuals were then ranked

within station on each characteristic on the basis of the average ratings. The average ratings. Ranks were converted to T-scores and trait-criterion correlations were comprrelations were computed on these data, pooling individuals across stations. Again, these results again, these results are shown in Table 4.

To check the appropriateness of pooling individuals across stations for the trait-criterion relationships, several analyses were run to determine the extent to determine the extent to which peers, supervisors, and entire station groups agreed with one another agreed with one another on the relative importance of the different traits. The latter determination the latter determination was assessed by the magnitude of the correlation which each trait had with each trait had with the criterion. That is, a high positive correlation was assumed to indicate importance of an attribute for the criterion judgment. For each analysis, trior each analysis, trait-criterion correlations were ranked within station and coefficient of concorcoefficient of concordance tests were then applied to all data. Using the peer criterion, supervisor criterion, supervisor criterion, and combined criterion with the traits, significant agreement wasificant agreement was obtained among the trait-criterion ranks for peer groups (W_c = .421, p < .05s (W_c = .421, p < .01), supervisor groups (W_c = .250, p < .05), and station groups (W_c = .371, p < oups (W_c = .371, p < .01) respectively. In summary, there was agreement across stations on the relatitations on the relative importance of the different characteristics for the criterion. Peers and supervisors furthermore agreed on the relative importance of the characteristics the characteristics as evidenced by a rank correlation (rho) of .84 (p < .01) between the orders o between the orders of correlation magnitude of traits with peer and supervisor criteria shown in or criteria shown in Table 4.

To the extent that peers and supervisors tend to agree on the relative value of different characteristics for overall performance, as illustrated agree on the relative ince, as illustrated in the preceding analysis, it seems reasonable to use the combined criterion combined criterion measure in future prediction studies. This criterion measure, as discussed easure, as discussed earlier, does have better discriminating potential than either the peer or either the peer or supervisor criterion alone when individuals are pooled across groups. across groups.

In an attempt to assess what structure might exist among the traits li among the traits listed in Table 4, the matrix of intercorrelations among those traits was submitte traits was submitted to a factor analysis. One additional item was included in the matrix, that be the matrix, that being a friendship-compatibility item derived from peer nominations of friends anations of friends and persons easy to get along with. Two factors accounted for 82 percent of th for 82 percent of the variance, one being a general factor and the other being a bipolar factor wig a bipolar factor with social-emotional attributes loading in one direction and task-oriented attrid task-oriented attributes in the other. The remaining factors were of value in the rotation pre in the rotation process. The rotated factor structure was such that three pairs of items emerged as of items emerged as meaningful concepts; the items in each pair had highly similar loadings on similar loadings on each of the rotated factors and each pair was somewhat different from the other rent from the other two. The three pairs or clusters were the following: (1) emotional control and otional control and acceptance of authority, (2) likability (from supervisor ratings) and frienor ratings) and friendship-compatibility (from peer nominations), and (3) industriousness and achlustriousness and achievement motivation. These clusters clearly represent the three facets of effthree facets of effective performance initially discussed in this paper, namely, emotional composure, emotional composure, social compatibility, and work motivation and effort.

The next procedure was to determine the relationship between the threeip between the three clusters extracted from factor analysis and the combined criterion measure, criterion measure, mainly in an effort to see whether or not such clusters could account for m could account for more variance in the criterion than could the individual items alone. To obtainms alone. To obtain cluster scores, the individual's T-scores for the two items in each clustertems in each cluster were averaged; the average T-scores were not re-ranked but maintained as they waintained as they were for subsequent analyses. Pooling individuals across stations, the cluster ations, the cluster scores were correlated with one another and with the combined criterion meambined criterion measure. These data are presented in Table 5.

TABLE 5

Trait Cluster-Criterion Correlations[a]

Trait Cluster	(E)	(S)	(T)	Criterio
Emotional composure (E)	(.72)	.55	.52	.70
Social Compatibility (S)		(.57)	.35	.80
Task motivation (T)			(.66)	.58

[a] Pearson r's based on T-scores; values in the diagonal refer to reliability estimates for the clusters based upon median levels of rater agreement on the combined items in each cluster; cluster (S) has a peer-supervisor agreement index. For N = 139, r = .16 (p < .05).

The clusters, individually, did not correlate much different criterion than did the individual items of emotional control, lik industriousness. On the other hand, although the clusters were o greater than the individual items in their correlation with the c were somewhat less correlated among themselves than the individua desirable condition if our goal is to find attributes which refle ponents of the criterion. Using the Wherry-Doolittle method, the correlation between all three clusters and the combined criterion highest multiple correlation obtained using various combinations items was .84.

The importance of the social compatibility cluster is appare tiple correlations are obtained between pairs of clusters and the By removing the social compatibility cluster, the value of R is r by combining the social compatibility cluster with the emotional task motivation cluster alone, the values of R are .86 and .85 re While both emotional composure and task motivation could contribu popularity, there are probably still other attributes of interper which increase a man's compatibility potential; qualities of cons and interpersonal warmth may not always be concomitant with contr triousness. It would seem to be these interpersonal qualities wh bined with composure and task motivation, make for effective perf small station.

One final analysis was that of determining the extent to whi clusters held up in their relationship with the criterion for dif groups and for military and civilian personnel. The values of R .87 for military and civilian personnel respectively. Comparing variations in beta weights occured and R values ranged from .76 t median value of R was .88. The three clusters, then, appeared to the same multiple relationship with the criterion for different g personnel.

In summary, the greatest amount of criterion variance was ac the three trait clusters of emotional composure, social compatibi motivation and effort. The addition of other characteristics to (i.e., attitudes towards job and project) did not improve the mul ship.

TABLE 5

Trait Cluster-Criterion Correlations[a]

Trait Cluster	(E)	(S)	(T)	Criterion
Emotional composure (E)	(.72)	.55	.52	.70
Social Compatibility (S)		(.57)	.35	.80
Task motivation (T)			(.66)	.58

[a] Pearson r's based on T-scores; values in the diagonal refer to reliability estimates for the clusters based upon median levels of rater agreement on the combined items in each cluster; cluster (S) has a peer-supervisor agreement index. For N = 139, r = .16 ($p < .05$).

The clusters, individually, did not correlate much differently with the criterion than did the individual items of emotional control, likability, and industriousness. On the other hand, although the clusters were only minutely greater than the individual items in their correlation with the criterion, they were somewhat less correlated among themselves than the individual items, a desirable condition if our goal is to find attributes which reflect unique components of the criterion. Using the Wherry-Doolittle method, the multiple correlation between all three clusters and the combined criterion was .89; the highest multiple correlation obtained using various combinations of individual items was .84.

The importance of the social compatibility cluster is apparent when multiple correlations are obtained between pairs of clusters and the criterion. By removing the social compatibility cluster, the value of R is reduced to .75; by combining the social compatibility cluster with the emotional composure or task motivation cluster alone, the values of R are .86 and .85 respectively. While both emotional composure and task motivation could contribute to a man's popularity, there are probably still other attributes of interpersonal skill which increase a man's compatibility potential; qualities of considerateness and interpersonal warmth may not always be concomitant with control and industriousness. It would seem to be these interpersonal qualities which, when combined with composure and task motivation, make for effective performance at the small station.

One final analysis was that of determining the extent to which the three clusters held up in their relationship with the criterion for different station groups and for military and civilian personnel. The values of R were .90 and .87 for military and civilian personnel respectively. Comparing station groups, variations in beta weights occured and R values ranged from .76 to .93, but the median value of R was .88. The three clusters, then, appeared to have about the same multiple relationship with the criterion for different groups of personnel.

In summary, the greatest amount of criterion variance was accounted for by the three trait clusters of emotional composure, social compatibility, and task motivation and effort. The addition of other characteristics to these clusters (i.e., attitudes towards job and project) did not improve the multiple relationship.

CONCLUSIONS

The present study was undertaken with the purpose of developing a criterion measure of effective individual performance for personnel who winter-over at small Antarctic stations. The best single criterion to date is a standard score based upon the combination of peer and supervisor choices of individuals with whom they would most prefer to return for further small station duty were they to return to the Antarctic. The standard score is based upon an average T-score (Mean = 50, SD = 10) derived from rank data. The criterion score can be quickly derived and can be used appropriately when personnel from different stations and years are pooled. Considering the nature of the criterion, its reliability appears to be adequate.

The specific qualities of behavior considered by the authors to have greatest face validity as attributes of effective performance were emotional composure, ability to get along with others, and work efforts. Through a series of analyses performed on data obtained from supervisor and peer evaluations, three behavior characteristic clusters were derived which represented the aforementioned three attributes of effective performance and which, in a multiple relationship, contributed the greatest amount of variance to the criterion measure of all evaluations available. The three clusters were of the following combinations of characteristics: emotional control and accepting authority (emotional composure), supervisor and peer estimates of likability (social compatibility), and industriousness and achievement motivation (task motivation).

The attribute of work effort and motivation appears to be more discriminating than work competency, at least in so far as can be determined by peer description data. This may very well be due, however, to an initial assignment process whereby only technically qualified persons are sent to small stations for duty; thus, there may be little if any variance in such an attribute among most wintering-over personnel. As for social compatibility, it appears that attributes such as considerateness (when in fact confronted with an interpersonal situation) may be more critical than one's social orientation (outgoing or retiring). Emotional control and acceptance of authority must be demonstrated to a degree sufficient to avoid interpersonal conflict, but an overly controlled individual may be somewhat inhibited in terms of task initiative and industriousness; there is a suggestion of this particularly with regard to acceptance of authority.

The primary value of the behavioral clusters just described lies in the fact that they are helpful in providing a conceptual definition of the criterion. This is particularly true when we introduce the problem of prediction--logically the next step in the present assessment program. It might now become more possible to construct or identify items (including clinical judgments) which on a conceptual or an empirical basis are more highly related to one or two of the behavior cluster constructs than to the criterion measure itself. The hope would be, then, to construct a battery of such items from screening data in such a way that greater predictive power with the criterion will be realized.

One might certainly argue, as has often been done, that regardless of the predictive efficiency achieved, there will always be the anchor men in any field group so long as our judgments of performance are based upon interpersonal comparisons. While this may be so, we might at least consider the possiblity of someday having to assemble a group of men on the basis of their individual potential effectiveness in a closed-group environment. Some familiarity with the types of attributes required, and what predictors are available, should be achieved by such a time. This problem, of course, also involves consideration of group composition variables, a problem of matching and contrasting

individuals on the basis of somewhat more idiosyncratic characteristics; this field of research is at present in its infancy.

REFERENCES

Anderhalter, O. F., Wilkins, W. L., & Rigby, Marilyn K. Peer ratings. St. Louis University, St. Louis, Mo.: ONR Contr. N7onr-40802 (NR151-092), Technical Report No. 2, Nov. 1952.

Bershire, J. R., & Nelson, P. D. Leadership peer ratings related to subsequent proficiency in training and in the fleet. U. S. Naval School of Aviation Medicine, Pensacola, Fla.: Special Report 58-20, Aug. 1958.

Fiske, D. W., & Cox, J. A. The consistency of ratings by peers. J. appl. Psychol., 1960, 44, 1, 11-17.

Gunderson, E. K. E., & Nelson, P. D. Adjustment criteria in Antarctica. U. S. Navy Med. NP Res. Unit, San Diego, Calif.: Report No. 62-1, Mar. 1962(a).

Gunderson, E. K. E., & Nelson, P. D. Attitude changes in small groups under prolonged isolation. U. S. Navy Med. NP Res. Unit, San Diego, Calif.: Report No. 62-2, Mar. 1962(b).

Hollander, E. P. Peer nominations on leadership as predictors of the pass-fail criterion in Naval Air Training. J. appl. Psychol., 1954, 38, 3, 150-153.

Nardini, J. E., Herrmann, R. S., & Rasmussen, J. E. Navy psychiatric assessment program in the Antarctic. Amer. J. Psychiat., Aug. 1963, 97-105.

Nelson, P. D., & Gunderson, E. K. E. Analysis of adjustment dimensions in small confined groups. U. S. Navy Med. NP Res. Unit, San Diego.: Report No. 62-3, Mar. 1962.

Nelson, P. D. Human adaptation to Antarctic station life. U. S. Navy Med. NP Res. Unit, San Diego, Calif.: Report No. 62-12, June 1962.

Nelson, P. D. A study of the validity of mail questionnaire data. U. S. Navy Med. NP Res. Unit, San Diego, Calif.: Report No. 63-7, Mar. 1963.

Wherry, R. J., & Fryer, D. Buddy ratings: popularity contest or leadership criteria? Sociometry, 1949, 12, 179-190.